Empirical-Likelihood-Based Inferences for Generalized Partially Linear Models

Hua Liang,
University of Rochester

Yongsong Qin,
Guangxi Normal University

Xinyu Zhang, and
Chinese Academy of Sciences

David Ruppert
Cornell University

Abstract

This paper considers generalized partially linear models. We propose empirical likelihood based statistics to construct confidence regions for the parametric and nonparametric componenets. The resulting statistics are shown to be asymptotically chi-squared distributed. Finite sample performance of the proposed statistics is assessed by simulation experiments. The proposed methods are applied to a dataset from an AIDS clinical trial.

Keywords

Confidence region; generalized additive models; least favorable curve; local linear regression; semiparametric estimation

1 Introduction

Generalized partially linear models (GPLM), a generalization of partially linear models to possibly non-Gaussian responses, assume that the conditional expectation of the response variables given the covariates can be represented as

$$E(Y|X,T)=\mu\{X'\beta+\theta(T)\}, \quad Var(Y|X,T)=\sigma^2 V(\mu), \tag{1}$$

where $\mu = \mu\{X'\beta + \theta(T)\}$, $\mu(\cdot)$ is a known link function, $V(\cdot)$ is a known function, β is an unknown $p \times 1$ vector, $T \in R^q$, θ is an unknown smooth function, and σ^2 is an unknown scalar parameter. Assume that T takes values in $\mathcal{T}$ a closed rectangle in R^q. We assume that (Y_i, X_i, T_i), $i = 1, 2, \cdots, n$, are independent and identically distributed data from model (1). GPLMs allow easier interpretation of the effect of each variables and are preferable to general nonparametric models (Stone, 1980) since they provides a partial remedy to the "*curse of dimensionality*," especially when q is small as is often the case. GPLMs are more flexible than the standard GLM because they combine both parametric and nonparametric

Hua Liang, PhD, Associate Professor, Department of Biostatistics and Computational Biology, University of Rochester Medical Center, Rochester, New York 14642, USA, hliang@bst.rochester.edu.

components when it is believed that the $E(Y|X, T)$ depends on variable X in a linear way but is nonlinearly related to other independent variables, T.

A special class of the GPLMs, partially linear models, have been intensively studied in literature. See for example, Engle et al. (1986), Speckman, (1988), Härdle, Liang & Gao (2000) and references therein. For GPLMs, Severini & Staniswalis (1994) applied the quasilikelihood principle proposed by Severini & Wong (1992), and Carroll et al. (1997) proposed two different estimation algorithms based on quasilikelihood and local kernel methods. Related topics have recently been studied by Lin & Carroll (2001) for longitudinal data, and Liang & Ren (2005) for measurement errors. It is worth pointing out that the quasilikelihood approach for the GPLM is different from the kernel-based smoothing method for partially linear models. The latter is simple and noniterative because the closed form of the estimators is available, while the former needs an iterative algorithm and an undersmoothing bandwidth.

Under mild regularity conditions, Severini & Staniswalis (1994) derived the asymptotics for the estimators of β and $\theta(t)$ that they proposed. In principle, these asymptotic results can be used to construct asymptotically correct confidence intervals of the parameters and pointwise confidence intervals for the nonparametric function. The finite-sample performance of the resulting confidence intervals may not be appealing because the complex structure of the covariance matrix, which needs to be estimated with estimates plugged-in for several parameters. In this paper, we propose an alternative for constructing regions for β and $\theta(t)$ using the empirical likelihood principle, which was originally studied by Hartley & Rao (1968) for sample surveys and by Thomas & Grunkemeier (1975) for survival analysis. Owen (2001) gave a comprehensive survey for empirical likelihood methods and related topics. The empirical likelihood method has many advantages over its competitors such as the normal-approximation-based method and the bootstrap method (see Hall & La Scala, 1990). These advantages include improvement of the confidence region, increase of accuracy of coverage because of using auxiliary information, easy implementation, avoiding estimating variances, and studentising automatically. Because of these features, the applications of empirical likelihood in parametric and nonparametric models have received a great amount of attention.

More recently, empirical likelihood based inference has been developed for semiparametric models, e.g., by Zhu & Xue (2006) who developed empirical likelihood confidence regions for the parameters of partially linear single-indexmodels. However, most research on empirical likelihood inference for semiparametric models has focused on the finite-dimensional parameter and has assumed a continuous response variables. In this paper, we study empirical likelihood inference for both the finite-dimensional parameters and the nonparametric functions in semiparametric models and we allow the response variable to be discrete. Our procedure is a generalization of empirical likelihood procedure to a combination of generalized linear models and nonparametric regression. This generalization is by no means straightforward. In Section 2, we will define the empirical likelihood ratio statistics for β and $\theta(t)$, derive the asymptotic distributions of the resulting empirical likelihood statistics, and explain how to establish the corresponding CI. In Section 3 we report the results of a simulation experiment to explore the finite sample performance of the proposed confidence intervals. The proposed methods will used to analyze a real dataset in Section 4. Section 5 gives a discussion. All technical derivations are given in the Appendix.

2 Empirical Likelihood Methods

Several authors have applied empirical likelihood to partially linear models, a special case of the GPLM. For example, Shi & Lau (1999) proposed an empirical likelihood based

confidence interval for the parameters of a partially linear model. Qin & Jing (2001) and Wang & Li (2002) considered the case in which the response variables Y_i are random censored. These authors proposed an empirical likelihood ratio for β and derived its asymptotic distribution, which is a sum of independent chi-squared distributions with unknown weights.

We first review briefly the quasi-likelihood estimators of β and $\theta(t)$ proposed by Severini & Staniswalis (1994). Denote the quasi-likelihood function by

$$Q(\mu, y) = \int_\mu^y \frac{s - y}{V(s)} ds.$$

Under some regularity conditions, $\sum_i Q(\mu, Y_i)$ behaves like a log-likelihood function for μ based on $Y_1, \cdots, Y_n$ and $Q(\mu, y)$ behaves like the logarithm of a density function for Y. Let K denote a kernel on R^q, and $h = h_n$ denote a sequence of bandwidths. For each fixed t and β, let $\hat{\theta}_\beta(t)$ denote the solution in η of

$$\sum_{i=1}^n K\left(\frac{t - T_i}{h}\right) \frac{\partial}{\partial \eta} Q\{\mu(\eta + X_i'\beta), Y_i\} = 0. \tag{2}$$

Let $\mathcal{T}_0$ denote a compact subset of int(T) and let $I_i = 1$ if $T_i \in \mathcal{T}_0$ and 0 otherwise. Given the estimator $\hat{\theta}_\beta(t)$, an estimator of β, $\hat{\beta}$ is then obtained by solving

$$\sum_{i=1}^n I_i \frac{\partial}{\partial \beta} Q\{\mu(\widehat{\theta_\beta}(T_i) + X_i'\beta), Y_i\} = 0. \tag{3}$$

The quasi-likelihood estimator of $\theta(t)$ is given by $\hat{\theta}_{\hat{\beta}}(t)$. The trimming by I_i of data near the boundary is employed to reduce boundary bias, which, for kernel regression estimators, can be quite serious and converges to zero at a slower rate than in the interior. In the univariate case, when $q = 1$, either a boundary-corrected kernel estimator or locally linear kernel estimator may be used instead. Although either of these methods may be extended to the multivariate case, the resulting technical details for the development of the asymptotic theory become cumbersome. For ease of notation, we present our results for the case $q = 1$ in the remainder of this paper.

2.1 Confidence region for β

Let β_0 denote the true value of β. Write

$$\omega_1\{\beta, \widehat{\theta_\beta}(T_i), Y_i, X_i, T_i\} = I_i \frac{\partial}{\partial \beta} Q[\mu\{\widehat{\theta_\beta}(T_i) + X_i'\beta\}, Y_i].$$

Based on the estimating equation (3) for β, we propose the empirical likelihood ratio statistic for β as follows.

$$\ell_1(\beta) = -2 \sup_{p_1, \cdots, p_n} \sum_{i=1}^n \log(np_i),$$

where p_i, $i = 1, \cdots, n$, are nonnegative numbers which satisfy

$$\sum_{i=1}^{n} p_i = 1, \quad \sum_{i=1}^{n} p_i \omega_1\{\beta, \widehat{\theta}_\beta(T_i), Y_i, X_i, T_i\} = 0.$$

By the Lagrange multiplier method, it can be shown that

$$\ell_1(\beta) = 2 \sum_{i=1}^{n} \log[\, 1 + \lambda_1' \omega_1\{\beta, \widehat{\theta}_\beta(T_i), Y_i, X_i, T_i\}]$$

where λ_1 is determined by

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\omega_1\{\beta, \widehat{\theta}_\beta(T_i), Y_i, X_i, T_i\}}{1 + \lambda_1' \omega_1\{\beta, \widehat{\theta}_\beta(T_i), Y_i, X_i, T_i\}} = 0 \text{ for } \beta \text{ in a neighbourhood of } \beta_0.$$

(4)

The asymptotic distribution of the empirical likelihood ratio statistic $\ell_1(\beta_0)$ is established in Theorem 2.1. Its proof is given in the Appendix.

Theorem 2.1—*Suppose that $nh^4 \to 0$ and the conditions (a)–(e) in the Appendix are satisfied. Then, as $n \to \infty$,*

$$\ell_1(\beta_0) \xrightarrow{d} \chi_p^2,$$

where β_0 is the true parameter value and χ_p^2 is a chi-square distributed random variable with p degrees of freedom.

Therefore, $CI_\beta = \{\beta | \ell_1(\beta) \le c_\alpha\}$ is a $1 - \alpha$ confidence region for β_0 where c_α satisfies $P(\chi_p^2 \le c_\alpha) = 1 - \alpha$.

2.2 Pointwise confidence region for θ(*t*)

Let $\eta = \theta(t)$ for fixed $t \in \mathscr{T}_0$, and $\hat{\beta}$ be a $\sqrt{n}$−consistent estimator of β_0. Denote

$$\omega_2(\eta, \widehat{\beta}, Y_i, X_i, T_i) = K\left(\frac{t - T_i}{h}\right) \frac{\partial}{\partial \eta} Q\{\mu(\eta + X_i' \widehat{\beta}), Y_i\},$$

where $K(\cdot)$ is a kernel function and h is a bandwidth. Based on the estimating equation (2) for η, we propose the empirical likelihood ratio statistic for η:

$$\ell_2(\eta) = -2 \sup_{p_1, \cdots, p_n} \sum_{i=1}^{n} \log(np_i),$$

where p_i, $i = 1, \cdots, n$, are nonnegative numbers which satisfy

$$\sum_{i=1}^{n} p_i = 1, \ \sum_{i=1}^{n} p_i \omega_2(\eta, \widehat{\beta}, Y_i, X_i, T_i) = 0.$$

A direct calculation implies that

$$\ell_2(\eta) = 2 \sum_{i=1}^{n} \log\{1 + \lambda_2 \omega_2(\eta, \widehat{\beta}, Y_i, X_i, T_i)\},$$

where λ_2 is determined by

$$\frac{1}{nh} \sum_{i=1}^{n} \frac{\omega_2(\eta, \widehat{\beta}, Y_i, X_i, T_i)}{1 + \lambda_2 \omega_2(\eta, \widehat{\beta}, Y_i, X_i, T_i)} = 0 \text{ for } \eta \text{ in a neighbourhood of } \eta(t).$$

(5)

The asymptotic distribution of the empirical likelihood ratio statistic $\ell_2(\eta)$ is given in Theorem 2.2. Its proof is given in the Appendix.

Theorem 2.2—*Suppose that $nh^4 \to 0$ and that the conditions (a)–(e) in the Appendix are satisfied. Then, as $n \to \infty$,*

$$\ell_2(\eta) \xrightarrow{d} \chi_1^2.$$

From Theorem 2.2, the confidence region for η with coverage probability $1 - \alpha (0 < \alpha < 1)$ can be constructed by $CI_\eta = \{\eta | \ell_2(\eta) \leq c_\alpha\}$, where c_α satisfies $P(\chi_1^2 \leq c_\alpha) = 1 - \alpha$.

Remark 1: *Theorems 2.1 and 2.2 indicate that undersmoothing is still needed as for the normal approximation theory (Carroll et al., 1997). To meet this requirement, we use existing bandwidth selection techniques to obtain the optimal bandwidth, $\hat{h}_{opt}$, An ad hoc bandwidth is generated by $\hat{h}_{opt} \times n^{-1/20} \log^{-1/5} n$, which ensures the bandwidth has correct order required in Theorems 2.1 and 2.2.*

3 Simulations

To illustrate the numerical performance for the proposed method, we conducted a small simulation experiment in which $n = 80, 100, 120$. We generated data from a logistic model

$$\text{logit}\{Y_i = 1 | X_i, T_i\} = X_i \beta + \theta(T_i),$$

where X_i is independent uniform $(-0.5, 0.5)$ component and T_i is uniformly distributed on $(0, 2)$. The parameter β is equal to 1, and the nonparametric function is $\theta(z) = \sin\{(z - a)/(b - a)\pi\}$ with $a = \sqrt{3}/2 - 1.645/\sqrt{12}$ and $b = \sqrt{3}/2 + 1.645/\sqrt{12}$.

In our nonparametric estimation implementation, to save computational time, we tried the simple bandwidth $h = an^{-1/4}(\log n)^{-1/5}$ for $a = 0.75, 1, 1.25, 1.5, 2$, which satisfy the condition in Theorems 1 and 2. We finally selected bandwidth via $h = 1.5n^{-1/4}(\log n)^{-1/5}$. The numerical results are fairly stable against shifting values of the selected bandwidth. We used the quartic kernel, $K(u) = 15/16(1 - u^2)^2 I_{(|u| \leq 1)}$. We generated 200 data sets in each

configuration. The empirical likelihood-based and normal approximation based confidence intervals for β are reported in Table 1. The lower and upper values are the averages of 200 simulated lower and upper values. The columns "AL" give the average length of the confidence intervals, while the column "CP(%)" gives the corresponding coverage probabilities of the 200 simulated datasets. The pointwise confidence intervals for the nonparametric function θ(t) at the selected four points t = 0.3, 0.8, 1.5 and 1.9 are presented in Table 2. A referee has asked us how the confidence intervals proposed compare to bootstrap confidence intervals, for which we used the naive bootstrap, i.e., resampled (X, Y, T), for 500 times. We provided the results for β in Table 1 and for the nonparametric component θ(t) in Table 2. These results basically coincide with those the results based on the normal approximation, and slightly deviate from those based on the proposed method in this paper. But the bootstrap implementation took a significantly longer amount of time compared to the empirical likelihood method. From the tables we may conclude that the coverage probabilities based on empirical likelihood method are mostly closer to the nominal level than those based on the normal approximation method, while the lengths of the empirical likelihood based intervals are slightly shorter than those based on the normal approximation method. The length of the estimated confidence intervals of β decreases with the increase of sample size.

4 Real Data Analysis

In recent years, one of the areas focused upon by AIDS researchers has been the relationship between viral load and CD4+ cell counts (Liang, Wu & Carroll, 2003; Liang, et al., 2004). This relationship is used to investigate the concordance and discordance between virologic and immunologic variables, which may help clinicians more deeply understand AIDS pathogenesis and improve therapy. Although antiretroviral therapy for HIV-1 infected patients has greatly improved in recent years, and administration of drug cocktails consisting of three or more drugs can reduce and maintain the viral load below the detection limit in many patients, it is unlikely that any combination of therapies can eradicate HIV in infected patients because of the existence of long-lived infected cells and sites within the body where drugs may not be effective. With the success of highly active antiretroviral therapy (HAART) against HIV infection, viral load (measured as viral RNA copies/mL) is suppressed and maintained at magnitudes that are below the limit of quantification, and the infection is considered chronic. Clinicians and patients are therefore nowadays more interested in achieving a viral load that is below the detection limit and in monitoring the immunologic system (measured by CD4+ cell counts).

In this section we analyze a dataset from the AIDS study PACTG 345 (Scott et al., 2001). Let Y be the indicator of a undetectable viral load level, let X be the CD4 cell count, and let T be the treatment time. In this study, 33 patients were enrolled as cohort II. Specimens were obtained on days 0, 1, 3, 7, 14, 28, 56, then irregularly through to the day 1155. A total of 559 HID-1 RNA measurements were obtained with 256 of these below the detection limit of 400 copies/mL. Thus, 45% of the viral loads were observed to be suppressed below the detection limit. Figure 1 presents the individual observations of plasma HID RNA concentration (viral load) after initial antiretroviral treatments. A main objective of the treatment is to suppress the viral load below the limit of detection.

We are interested in the relationship between the binary viral load measurement and CD4+ cell counts. A parsimonious model of this relationship is biologically and clinically important because these variables are good biomarkers for anti-HIV treatment and may be used to evaluate antiretroviral therapies. An obvious model is logistic regression, with X and T having linear effects on the logit scale, because it is easily implemented and interpreted. A concern, however, is whether this model can appropriately capture curvature in the effect of

T due to drug resistance or noncompliance. To address this concern, we used the method of Härdle, Mammen & Müller (1998) to check if a logistic model is appropriate, and obtained a *p*-value less than 10^{-4}, which reflects that the traditional logistic regression is not flexible enough to fit this data set well. We therefore used a partially logistic model, described in (6), to fit the dataset and use the proposed method to obtain the confidence intervals for parametric and nonparametric components.

$$\text{logit}\{E(Y|X,T)\}=X\beta+\theta(T),$$

(6)

where $\theta(t)$ is a unknown smooth function. The estimate of β is 0.216, the positive value of which reflects the increased chance of RNA below the detection limit at higher levels of CD4+ counts. The 95% confidence intervals for β based on the normal approximation and the proposed empirical likelihood methods are (−0.202, 0.634) and (0.081, 0.514). These two confidence intervals convey different messages. The former interval indicates that the chance of RNA below the detection limit is not statistically significantly related to CD4 cell count, but the latter interval yields an inverse impression. We prefer to the conclusion based on the empirical likelihood method according to biological meanings and the simulation performance of this method. The pointwise estimates of $\theta(t)$ and associated confidence regions based on these two methods are shown in Figure 2, in which the solid line is the estimated pattern of $\theta(t)$, the dotted lines and broken lines are the confidence regions based on empirical likelihood and normal approximation methods. The former gives a narrower region than the latter.

5 Discussion

To simply inference for GPLMs, we proposed an empirical likelihood-based approach to constructing confidence regions for β and $\theta(t)$. The proposed approach is remarkably simpler than its counterpart based on the asymptotic normality of quasilikelihood estimators (Severini & Staniswalis, 1994) and easily executable. The finite-sample performance of the proposed statistics shows promise. In this article, we used local linear regression when we handled nonparametric function $\theta(t)$. There are many different alternatives to the local constant kernel regression in (2), including higher degree local polynomial kernel methods, smoothing splines, and regression splines. The details for these methods need further investigation in our setting. We chose the constant kernel regression because theoretical results can be derived (Severini & Staniswalis, 1994).

Model (1) may be extended to a generalized additive partially linear model in the form of

$$E(y_i|X_i,\mathbf{T}_i)=\mu\left\{X_i'\beta+\sum_{k=1}^{K}\theta_k(T_{k,i})\right\},$$

where $\mathbf{T}_i=(T_{1,i},\ldots,T_{K,i})'$ is a *K*-dimensional vector. The study of this model is interesting and requires additional efforts, but it is beyond the scope of this paper.

Acknowledgments

References

Carroll RJ, Fan J, Gijbels I, Wand MP. Generalized partially linear single-index models. J. Am. Statist. Assoc 1997;92:477–489.

Chen SX, Qin YS. Empirical likelihood confidence intervals for local linear smoothers. Biometrika 2000;87:946–953.

Engle RF, Granger CWJ, Rice J, Weiss A. Semiparametric estimates of the relation between weather and electricity sales. J. Am. Statist. Assoc 1986;81:310–320.

Hall P, La Scala B. Methodology and algorithms of empirical likelihood. Int. Statist. Rev 1990;58:109–127.

Hartley HO, Rao JNK. A new estimation theory for sample surveys. Biometrika 1968;55:547–557.

Härdle W, Mammen E, Müller M. Testing parametric versus semiparametric modeling in generalized linear models. J. Am. Statist. Assoc 1998;93:1461–1474.

Härdle, W.; Liang, H.; Gao, J. Partially Linear Models. Heidelberg: Springer Physica-Verlag; 2000.

Liang H, Ren HB. Generalized partially linear measurement error models. J. Comp. Graph. Statist 2005;14:237–250.

Liang H, Wu HL, Carroll RJ. The relationship between virologic and immunologic responses in AIDS clinical research using mixed-effect varying-coefficient semiparametric models with measurement error. Biostatistics 2003;4:297–312. [PubMed: 12925523]

Liang H, Wang S, Robins JM, Carroll RJ. Estimation in partially linear models with missing covariates. J. Am. Statist. Assoc 2004;99:357–367.

Lin XH, Carroll RJ. Semiparametric regression for clustered data using generalized estimating equations. J. Am. Statist. Assoc 2001;96:1045–1056.

Owen AB. Empirical likelihood ratio confidence regions. Ann. Statist 1990;18:90–120.

Owen, AB. Empirical likelihood. New York: Chapman and Hall; 2001.

Qin GS, Jing BY. Censored partial linear models and empirical likelihood. J. Mult. Anal 2001;78:37–61.

Qin J. Empirical likelihood ratio based confidence intervals for mixture proportions. Ann. Statist 1999;27:1368–1384.

Qin J, Lawless J. Empirical likelihood and general estimating equations. Ann. Statist 1994;22:300–325.

Scott ZA, Chadwick EG, Gibson LL, et al. Infrequent detection of HIV-1-specific, but not cytomegalovirus-specific, CD8+T cell responses in young HIV-1-infected infants. J. Immunology 2001;167:7134–7140. [PubMed: 11739536]

Severini TA, Staniswalis JG. Quasilikelihood estimation in semiparametric models. J. Am. Statist. Assoc 1994;89:501–511.

Severini TA, Wong WH. Profile likelihood and conditionally parametric models. Ann. Statist 1992;20:1768–1802.

Shi J, Lau TS. Empirical likelihood for partially linear models. J. Mult. Anal 1999;72:132–148.

Speckman P. Kernel smoothing in partial linear models. J. R. Statist. Soc. B 1988;50:413–436.

Stone CJ. Optimal rates of convergence for nonparametric estimators. Ann. Statist 1980;8:1348–1360.

Thomas DR, Grunkemeier GL. Confidence interval estimation of survival probabilities for censored data. J. Am. Statist. Assoc 1975;70:865–871.

Wang QH, Li G. Empirical likelihood semiparametric regression analysis under random censorship. J. Mult. Anal 2002;83:469–486.

Zhu LX, Xue LG. Empirical likelihood confidence regions in a partially linear single-index model. J. R. Statist. Soc. B 2006;68:549–570.

Appendix

Conditions

The following assumptions are standard in studies of GPLMs, and we assume these hold throughout the article. Write $\rho_1(u) = \{d\mu(u)/du\} \ V^{-1}\{\mu(u)\}$, and $q_1(u, y) = \{y-\mu(u)\}\rho_1(u)$.

a. The density function $f(t)$ of T is positive and continuous at the point $t_0 \in \mathcal{T}$

b. The function $\mu(u)$ is twice differentiable in u.

c. The function $\theta^{(2)}(t)$ is continuous at the point $t_0 \in \mathcal{T}$

d. With $R=\theta(T)+X'\beta$, $E\{q_1^2(R, Y)|T=t\}$, $E\{q_1^2(R, Y)X|T=t\}$, and $E\{q_1^2(R, Y)XX'|T=t\}$ are twice differentiable in t.

e. $E\{q_1^{2+\delta}(R, Y)\}<\infty$, for some $\delta > 2$.

Proof of Theorem 2.1

Denote AA' by $A^{\otimes 2}$, $\Lambda_{\beta_0,i}=\theta_{\beta_0}(T_i)+X_i'\beta_0$ and $\Lambda_i=\theta(T_i)+X_i'\beta_0$ for $i = 1, \ldots, n$. $\Xi_n = \max_{1\leq i\leq n} \|\omega_1\{\beta_0, \hat{\theta}_{\beta_0}(T_i), Y_i, X_i, T_i\}\|$. We first show that

$$n^{-1/2}\sum_{i=1}^{n}\omega_1\{\beta_0, \widehat{\theta}_{\beta_0}(T_i), Y_i, X_i, T_i\} \xrightarrow{d} N(0, G),$$

(7)

$$n^{-1}\sum_{i=1}^{n}[\omega_1\{\beta_0, \widehat{\theta}_{\beta_0}(T_i), Y_i, X_i, T_i\}]^{\otimes 2}=G+o_p(1),$$

(8)

and

$$\Xi_n=o(n^{1/2}), a.s.,$$

(9)

where Γ is a positive definite matrix in form of

$$\sigma^2 E\left[I_1\left\{X_1+\frac{\partial\theta_\beta(T_1)}{\partial\beta}\bigg|_{\beta=\beta_0}\right\}^{\otimes 2}\frac{\{\mu'(\Lambda_1)\}^2}{V\{\mu(\Lambda_1)\}}\right].$$

Recall that $Q(\mu, y)$ behaves like the logarithm of a density function for Y, and that $\theta_\beta(t)$ is a least favorable curve and thus proposition 2 of Severini &Wong (1992) holds, which are shown in the proof of proposition 1 of Severini and Staniswalis (1994). Accordingly, applying (2) in Section 6 of Severini & Wong (1992) (here our $\sum_{i=1}^{n}\omega_1\{\beta, \theta_\beta(T_i), Y_i, X_i, T_i\}$ corresponds to $\dfrac{dL_n(\theta, \lambda_\theta)}{d\theta}$ of Severini & Wong (1992)), we obtain

$$n^{-1/2}\sum_{i=1}^{n}\omega_1\{\beta_0, \widehat{\theta}_{\beta_0}(T_i), Y_i, X_i, T_i\}=n^{-1/2}\sum_{i=1}^{n}\omega_1\{\beta_0, \theta_{\beta_0}(T_i), Y_i, X_i, T_i\}+o_p(1).$$

(10)

Furthermore,

$$
\begin{aligned}
\sum_{i=1}^{n} \omega_1\{\beta_0, \theta_{\beta_0}(T_i), Y_i, X_i, T_i\} &= \sum_{i=1}^{n} I_i \left\{ X_i + \left.\frac{\partial \theta_\beta(T_i)}{\partial \beta}\right|_{\beta=\beta_0} \right\} \frac{\mu'(\Lambda_{\beta_0,i})}{V\{\mu(\Lambda_{\beta_0,i})\}} \{Y_i - \mu'(\Lambda_{\beta_0,i})\} \\
&= \sum_{i=1}^{n} I_i \left\{ X_i + \left.\frac{\partial \theta_\beta(T_i)}{\partial \beta}\right|_{\beta=\beta_0} \right\} \frac{\mu'(\Lambda_i)}{V\{\mu(\Lambda_i)\}} \{Y_i - \mu'(\Lambda_i)\}.
\end{aligned}
$$

(11)

(7) follows from (10), (11) and a central limit theorem. The proofs of (8) and (9) are trivial.

From (7), (8) and (9), and the arguments similar to the proof of (2.14) in Owen (1990), we can show that

$$
\lambda_1 = O_P(n^{-1/2}).
$$

(12)

Recall (4). It is readily seen by a direct calculation and (12) that

$$
\begin{aligned}
\lambda_1 = & \left(n^{-1} \sum_{i=1}^{n} [\omega_1\{\beta_0, \widehat{\theta}_{\beta_0}(T_i), Y_i, X_i, T_i\}]^{\otimes 2} \right)^{-1} \\
& \times \left[n^{-1} \sum_{i=1}^{n} \omega_1\{\beta_0, \widehat{\theta}_{\beta_0}(T_i), Y_i, X_i, T_i\} \right] + o_p(n^{-1/2}).
\end{aligned}
$$

Thus, using Taylor expansion, we have

$$
\begin{aligned}
\ell_1(\beta_0) = & \; n \left[\frac{1}{n} \sum_{i=1}^{n} \omega_1\{\beta_0, \widehat{\theta}_{\beta_0}(T_i), Y_i, X_i, T_i\} \right]' \\
& \times \left(\frac{1}{n} \sum_{i=1}^{n} [\omega_1\{\beta_0, \widehat{\theta}_{\beta_0}(T_i), Y_i, X_i, T_i\}]^{\otimes 2} \right)^{-1} \\
& \times \left[\frac{1}{n} \sum_{i=1}^{n} \omega_1\{\beta_0, \widehat{\theta}_{\beta_0}(T_i), Y_i, X_i, T_i\} \right] + o_p(1).
\end{aligned}
$$

The proof is complete from (7) and (8).

Proof of Theorem 2.2

Denote by $f_0(\cdot)$ the probability density function of T. Write

$$
H(Y, X, T) = \frac{\mu'\{\theta(T) + X'\beta_0\}}{V[\mu\{\theta(T) + X'\beta_0\}]} [Y - \mu'\{\theta(T) + X'\beta_0\}].
$$

We first show that

$$
(nh)^{-1/2} \sum_{i=1}^{n} \omega_2(\eta, \widehat{\beta}, Y_i, X_i, T_i) \xrightarrow{d} N(0, G_0),
$$

(13)

$$(nh)^{-1} \sum_{i=1}^{n} \omega_2^2(\eta, \widehat{\beta}, Y_i, X_i, T_i) = G_0 + o_p(1),$$

(14)

where $\Gamma_0 = \int K^2(u) du \cdot f_0(t) E\{H^2(Y, X, T) | T = t\}$.

From Taylor expansion and the fact that

$$\sqrt{n}(\widehat{\beta} - \beta_0) = O_p(1), \quad \frac{1}{nh} \sum_{i=1}^{n} \frac{\partial}{\partial \beta} \omega_2(\eta, \beta, Y_i, X_i, T_i) \bigg|_{\beta = \beta_0} = O_p(1),$$

it can been shown that

$$(nh)^{-1/2} \sum_{i=1}^{n} \omega_2(\eta, \widehat{\beta}, Y_i, X_i, T_i) = (nh)^{-1/2} \sum_{i=1}^{n} \omega_2(\eta, \beta_0, Y_i, X_i, T_i)$$
$$+ \sqrt{nh}(\widehat{\beta} - \beta_0)'(nh)^{-1} \sum_{i=1}^{n} \frac{\partial}{\partial \beta} \omega_2(\eta, \beta, Y_i, X_i, T_i) \bigg|_{\beta = \beta_0} \{1 + o_p(1)\}$$
$$= (nh)^{-1/2} \sum_{i=1}^{n} \omega_2(\eta, \beta_0, Y_i, X_i, T_i) + o_p(1).$$

(15)

Moreover,

$$\sum_{i=1}^{n} \omega_2(\eta, \beta_0, Y_i, X_i, T_i) = \sum_{i=1}^{n} K\left(\frac{t - T_i}{h}\right) \frac{\mu'(\Lambda_{\beta_0, i})}{V\{\mu(\Lambda_{\beta_0, i})\}} \{Y_i - \mu'(\Lambda_{\beta_0, i})\}$$
$$= \sum_{i=1}^{n} K\left(\frac{t - T_i}{h}\right) \frac{\mu'(\Lambda_i)}{V\{\mu(\Lambda_i)\}} \{Y_i - \mu'(\Lambda_i)\}.$$

(16)

Thus (13) follows from (15), (16) and a central limiting theorem. The proof of (14) is trivial.

Write

$$\bar{U}_j = (nh)^{-1} \sum_{i=1}^{n} \omega_2^j(\eta, \widehat{\beta}, Y_i, X_i, T_i), \text{ for } j = 1, 2.$$

From (15), using a central limiting theorem, it can be shown that
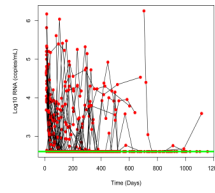
$$\bar{U}_1 = O_p\{(nh)^{-1/2} + h^2\}.$$

Combining with (5), (13) and (14), we have

$$\lambda_2 = (\bar{U}_2)^{-1} \bar{U}_1 + O_p[\{(nh)^{-1/2} + h^2\}^2].$$

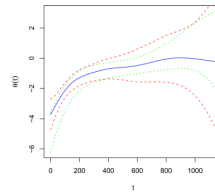Furthermore, by Taylor expansion, we obtain

$$
\begin{aligned}
\ell_2(\eta) &= 2\lambda_2 \sum_{i=1}^{n} \omega_2(\eta, \widehat{\beta}, Y_i, X_i, T_i) \\
&\quad - \lambda_2^2 \sum_{i=1}^{n} \omega_2^2(\eta, \widehat{\beta}, Y_i, X_i, T_i) + O_p[\{(nh)^{-1/2} + h^2\}^3] \\
&= (nh)\bar{U}_1^2(\bar{U}_2)^{-1} + O_p[\{(nh)^{-1/2} + h^2\}^3]
\end{aligned}
\tag{17}
$$

Theorem 2.2 follows from (13), (14), and (17).

Figure 1.
Viral load measurements of plasma HID RNA concentration in the PACTG 345 study. The detection limit of 400 copies of HID RNA per mL of plasma is indicated by the horizontal line.

Figure 2.
Confidence intervals of θ(*t*) obtained by using the normal approximation and empirical likelihood methods. The solid line is the estimated curve of θ(*t*), while the dotted and broken lines are the pointwise confidence intervals based on empirical likelihood and normal approximation methods.

Table 1

95% normal approximation (Normal), bootstrap (BC) and empirical likelihood (EL) based confidence intervals for $\beta = 1$ along with the averages of the left (LE) and right endpoints (RE) of the confidence intervals, their average lengths (AL), and the associated coverage probabilities (CP) for the simulated data.

n	Normal				BC				EL			
	CP(%)	LE	RE	AL	CP(%)	LE	RE	AL	CP(%)	LE	RE	AL
80	93.0	−0.568	2.889	3.457	97.0	−0.547	2.904	3.451	96.0	−0.604	2.823	3.427
100	97.0	−0.563	2.493	3.056	96.5	−0.541	2.415	2.955	95.0	−0.514	2.468	2.982
120	97.0	−0.352	2.384	2.736	97.0	−0.362	2.385	2.747	94.5	−0.295	2.398	2.692

Table 2

95% normal approximation (Normal), bootstrap (BC) and empirical likelihood (EL) based pointwise confidence intervals for nonparametric function θ(*t*) at the four selected points along with the averages of the left (LE) and right endpoints (RE) of the confidence intervals, their average lengths (AL), and the associated coverage probabilities (CP) for the simulated data.

t	*n*	Normal				BC				EL			
		CP(%)	LE	RE	AL	CP(%)	LE	RE	AL	CP(%)	LE	RE	AL
0.3	80	94.5	−1.091	0.793	1.885	95.0	−1.038	0.829	1.867	94.0	−0.841	0.914	1.756
	100	97.5	−0.995	0.586	1.581	97.0	−0.999	0.587	1.586	96.0	−0.893	0.589	1.482
	120	98.5	−0.844	0.418	1.262	96.5	−0.840	0.428	1.267	96.5	−0.823	0.442	1.265
0.8	80	94.7	−0.057	1.711	1.768	95.5	−0.044	1.671	1.715	95.5	−0.033	1.648	1.681
	100	96.5	0.144	1.613	1.469	97.0	0.133	1.614	1.482	95.5	0.093	1.613	1.519
	120	97.0	0.103	1.448	1.345	96.5	0.100	1.427	1.327	94.5	0.110	1.336	1.225
1.5	80	94.0	−1.064	0.431	1.495	94.5	−1.096	0.404	1.500	96.5	−1.032	0.417	1.449
	100	97.5	−1.070	0.278	1.348	97.0	−1.048	0.283	1.332	93.5	−0.786	0.303	1.090
	120	96.5	−1.037	0.236	1.273	97.0	−1.002	0.265	1.267	96.0	−0.965	0.259	1.223
1.9	80	96.0	−2.030	0.157	2.187	96.5	−1.974	0.172	2.146	96.0	−1.962	0.160	2.123
	100	97.0	−1.957	0.043	2.000	96.5	−1.973	0.030	2.003	95.5	−1.629	0.284	1.913
	120	97.5	−1.963	0.024	1.987	97.0	−1.953	0.032	1.985	95.5	−1.630	0.102	1.732